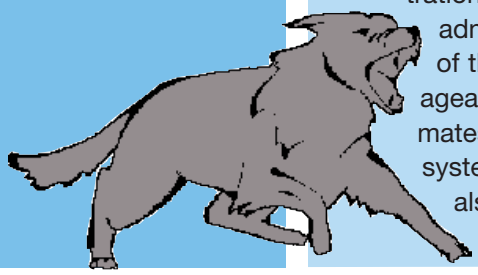
WAREWULF HELPS LBNL MANAGE CLUSTERS SO SCIENTISTS CAN FOCUS ON SCIENCE

In 2003, Berkeley Lab launched a new program called Scientific Cluster Support to increase the contribution of scientific computing to LBNL research projects by facilitating the use of Linux clusters. The program, run by the UNIX Systems Group in the Information Technologies and Services Division, helps selected research groups with choosing and procuring the right cluster configuration, setting up and configuring the system, helping researchers get their applications onto the cluster, then providing ongoing systems administration and cybersecurity support.



As Linux has evolved, the costs of entry into cluster HPC have been lowered, but setting up effective clusters — installing, configuring, and optimizing systems — is time consuming and costly. Linux clusters that facilitate scalable administration models are typically designed around ease-of-use or flexibility, but rarely accomplish both. Setting up a cluster using a generic cluster distribution system is a fairly straightforward proposition. The challenge is in customizing a cluster to meet the specific needs of computational scientists. Creating, managing and distributing a customized Linux system to an arbitrary number of nodes using a master-slave relationship has been a real challenge due to the fact that there is no tool that optimally scales the administration time, but still allows for extensive customization. Having such a tool would benefit both users and the support staff.

The Warewulf Solution

Enter Warewulf, the LBNL-developed tool that elegantly solves this problem, using readily available tools, and has become a stable solution for many production HPC clusters. Warewulf is a cluster implementation toolkit that facilitates the process of cluster installation and long-term administration. It does this by changing the administration paradigm to make all of the slave-node file systems manageable from one point, and automates the distribution of the node file system during node boot. Warewulf also creates a central administration model for all the slave nodes and includes the tools necessary to build configuration files, and monitor and control the nodes. The toolkit is totally customizable and can be adapted to just about any type of cluster. From the software administration perspective, it does not make any difference if you are running two nodes or 500 nodes, the procedure is the same. Also, because it uses a standard chroot-able file system for all of the nodes, it is extremely configurable and easily lends itself to custom environments.

While Warewulf was designed to enable HPC solutions, it is not an HPC solution in itself. Warewulf is more along the lines of a *distributed Linux distribution* — or more specifically, a system for replicating and managing small, lightweight Linux systems from one master. Using Warewulf, HPC packages such as LAM-MPI/MPICH, SGE, and PVM can be easily deployed throughout the cluster.

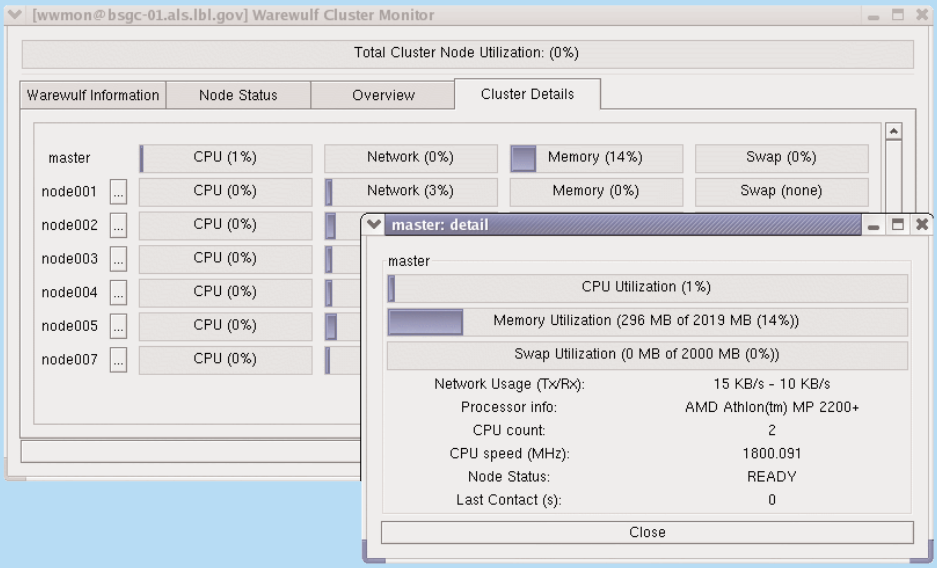
The Distributed Linux Distribution: Although it was designed with HPC in mind, Warewulf is just as flexible as a homegrown cluster, but also scales very well. As a result of this flexibility and ease of customization, Warewulf has been used not only on production HPC implementations, but also development systems like KASY0 — the first system to break the one hundred dollar-per-gigaflop/s barrier — and non-HPC solutions, such as web server cluster farms, intrusion detection clusters, and high-availability clusters.

SCS Scientific Cluster Support

TEID 9003

Supported under the SciDAC Program, DOE Office of Science at multiple sites

Keeping things simple: Instead of using a complex SQL back-end integrated into a complex front-end, Warewulf utilizes simple, well-known techniques to manage nodes. This means that there is less to break, and customization is simple and straightforward. Configuration and changes can be done via graphical utilities, text utilities, or the ASCII configuration files directly.



Easily integrated into standard RPM-based distributions: Warewulf is distributed not only as a tarball, but also a set of RPMS which facilitates installation onto a standard RPM-based master node. Warewulf does not modify any of the core packages on the master node; all updates and security fixes come from the distribution vendor, and are not re-distributed by Warewulf.

Some of the advantages of using Warewulf are:

- No installations on nodes. On every boot, the node receives its file system as a pristine looped file system.
- No version creep between nodes, as each node is guaranteed to have identical file system images at each boot.
- Troubleshooting node problems is simple. If one node is having a problem, it must be hardware-related — if other nodes are working, there is no other possibility.
- No node hard disk drives are needed (but they are supported as swap and data or scratch mount points).
- Changing software components or a file system on all nodes is as easy as changing one node.
- Supports many types of cluster nodes natively, even heterogeneous hardware of similiar architecture.
- Extensively customizable, and allows for many different cluster solutions.
- It is distribution neutral.
- Ease of scalability: allows one master to boot an arbitrary number of nodes.

Summary

As Linux has made purchasing a HPC cluster cost-effective, Warewulf has helped make clusters easy to build, customize, implement and maintain. Because of its centralized administration features and extensibility, Warewulf has become the standard cluster implementation tool for the Scientific Cluster Support project at Lawrence Berkeley National Laboratory, along with several other organizations. Warewulf is released by Berkeley Lab under a general public license (GPL), and community usage and development are encouraged.

References

The Warewulf home page: http://warewulf-cluster.org/

LBNL's Scientific Cluster Support project: http://scs.lbl.gov/

The Etherboot project: http://etherboot.sourceforge.net/

KASY0 at the University of Kentucky: http://aggregate.org/KASY0/

SCS Scientific Cluster Support